The EMBO Journal vol.6 no.1 pp.161–168, 1987

Contactpoint analysis of the HeLa nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix

Erik de Vries, Wim van Driel, Sander J.L.van den Heuvel and Peter C.van der Vliet

Laboratory for Physiological Chemistry, State University of Utrecht, Vondellaan 24a, 3521 GG Utrecht, The Netherlands

Communicated by H.S.Jansz

Nuclear factor I (NFI) is a HeLa sequence-specific DNA-binding protein that is required for initiation of adenovirus (Ad) DNA replication and may be involved in the expression of several cellular genes. The interaction between NFI and its binding site on the Ad2 origin has been studied. Methylation interference and protection, u.v. irradiation of 5-BrdU substituted DNA and ethylation interference revealed major groove contacts with G and T, and phosphate backbone contacts. Computer stereographics show that (i) the contacts are located in two blocks showing dyad symmetry to each other and (ii) 22 out of 23 contacts are accessible from one side of the helix. Inversion of the NFI binding site did not change the NFI dependent stimulation of Ad2 DNA replication in a reconstituted system. All data are compatible with NFI binding as a dimer at one side of the DNA helix.

Key words: adenovirus DNA replication/contactpoint analysis/ nuclear factor I/sequence-specific DNA-binding protein/ 5-bromo-deoxyuridine

Introduction

Nuclear factor I (NFI), a human sequence-specific DNA-binding protein, was first detected in HeLa cell nuclear extracts by virtue of its ability to enhance initiation of adenovirus DNA replication *in vitro* (Nagata *et al.*, 1982). NFI has a minimal recognition sequence of 15 or 16 bp, located internally in the origin region, 25 bp removed from the actual replication start site at the end of the linear genome (Leegwater *et al.*, 1985). Binding to this site (K_D is 2.1×10^{-11} M, Rosenfeld and Kelly, 1986) is essential for stimulation of replication *in vitro* (Rawlins *et al.*, 1984; Guggenheimer *et al.*, 1984; Leegwater *et al.*, 1985; De Vries *et al.*, 1985) whereas *in vivo* this sequence is also required for efficient replication (Hay, 1985). A number of cellular NFI binding sites have been identified. These include randomly isolated sites (Gronostajski *et al.*, 1984) as well as sites located upstream the c-*myc*, IgM and chicken lysozyme genes (Siebenlist *et al.*, 1984; Hennighausen *et al.*, 1985; Borgmeyer *et al.*, 1984). In most cases studied the binding sites coincide with a DNase I hypersensitive site in chromatin. A protein isolated from chicken cells, the TGGCA binding protein (Borgmeyer *et al.*, 1984) which is implicated in tissue specific lysozyme gene expression (Sippel *et al.*, 1986) is functionally homologous to NFI and has the same template requirements for efficient binding (Leegwater *et al.*, 1986). These results are in agreement with a role of NFI in gene expression, but firm evidence for such a function in cellular metabolism is still lacking.

Comparison of the NF1 binding sites of different adenovirus serotypes and cellular binding sites results in a consensus sequence

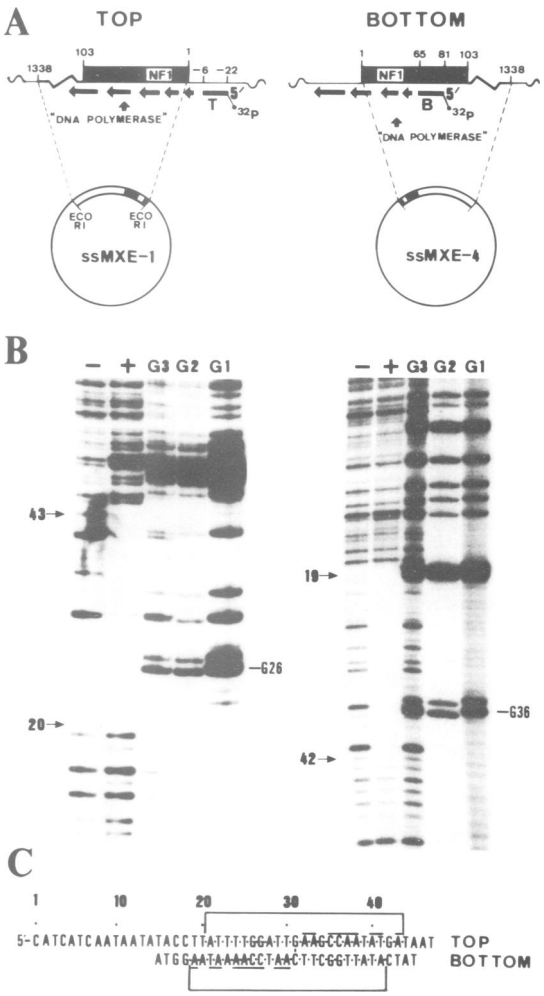


Fig. 1. Symmetrical DNase I footprint of the NFI binding site. (**A**) Schematic procedure for the synthesis of labeled binding sites. Synthetic oligonucleotides T (top) and B (bottom) are phosphorylated with [γ-^{32}P]ATP and T4 polynucleotide kinase to more than 90% completion. Hybridization to complementary single-stranded M13 clones MXE-2 and MXE-4, containing the Ad2 1-1338 insert, followed by primer extension with DNA polymerase allows synthesis of a double-stranded NFI site 5′-end-labeled at either the top or the bottom strand. The left end inverted terminal repeat of Ad2 (103 bp), containing the origin of replication including the NFI binding site, is indicated as a black bar. (**B**) DNAse I digestion was performed in the absence (−) or the presence (+) of purified NFI. The products were analyzed by electrophoresis on an 8% polyacrylamide gel followed by autoradiography. Arrows mark the outermost protected phosphate bonds. Bond 20 is the bond at the 3′ side of nucleotide 20 and so on. **Lane G_2** contains a chemical sequencing G-reaction of the same fragment. In **lane G_1** the products from lane G_2 are incubated with T4 nucleotide kinase to remove the 3′-PO_4 groups (Cameron and Uhlenbeck, 1977) to compare directly with the DNase I products that also lack 3′-PO_4 groups. **Lane G_3** contains a mixture of the products of marker lane G_2 and a DNase I digest in the absence of NFI as an internal control. (**C**) Sequence of the NFI binding site on the Ad2 origin of replication. P-bonds within the brackets are protected against DNase I. Numbering starts in the first base pair of the left end of Ad2. Palindromic squences are underlined and the axis of symmetry through base pair 31 is indicated.

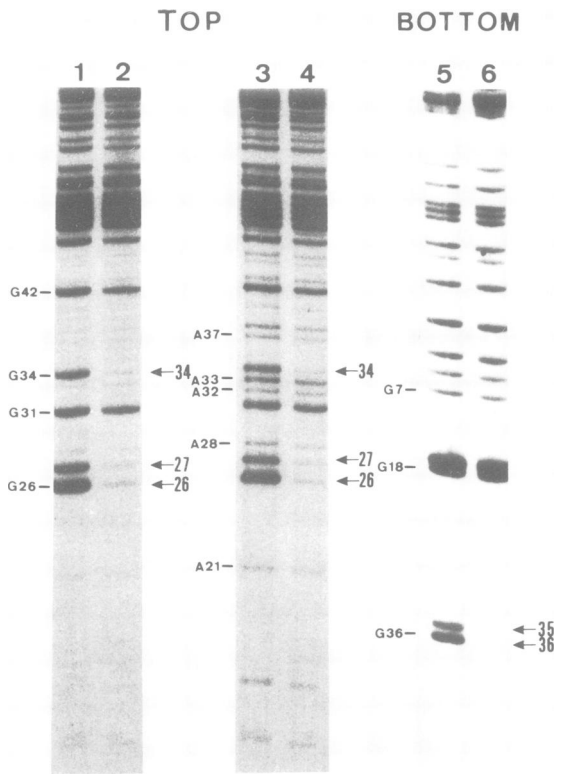


Fig. 2. Methylation protection by NFI. The 454 bp *Eco*RI−*Pvu*II fragment of XD7, 5′-end-labelled at the *Eco*RI site, was used. For the bottom strand we made use of the 270 bp *Hin*dIII−*Pvu*II fragment of pPL73, 5′-end-labeled at the *Hin*dIII site. Dimethylsulphate modification was carried out in the presence (**lanes 2, 4** and **6**) or absence (**lanes 1, 3** and **5**) of NFI, and cleaved next to G's (**lanes 1, 2, 5** and **6**) or G's + A's (**lanes 3** and **4**). Samples were analyzed on a 10% polyacrylamide gel. G's that are protected against methylation are indicated by arrows. Numbering of nucleotides is as indicated in Figure 1.

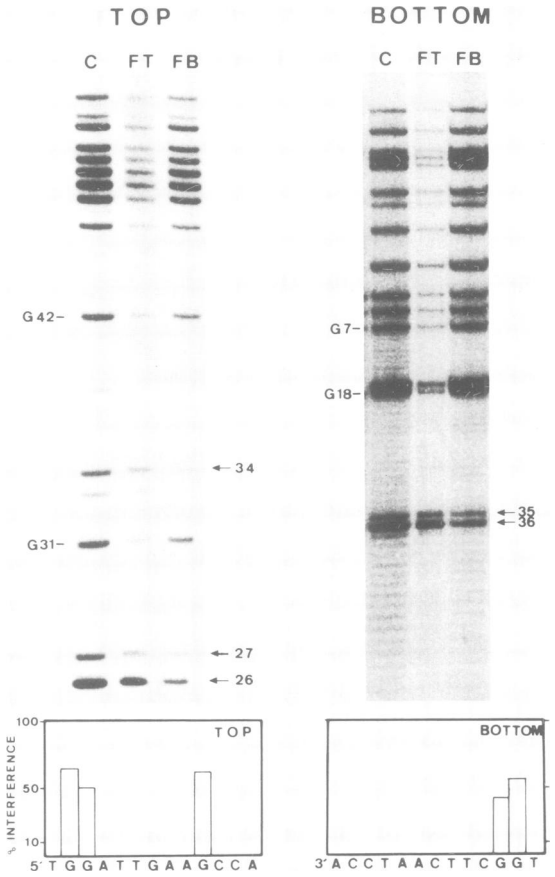


Fig. 3. Methylation interference of NFI-binding. The DNA fragments as described in the legend to Figure 2 were used. Partially methylated fragments were incubated with NFI and separated by filtration over nitrocellulose in a protein bound (Filterbound, FB) and an unbound (Filtrate, FT) fraction and analyzed on an 8% polyacrylamide gel. For comparison unseparated DNA is shown (C). G's at which methylation interferes with binding of NFI are indicated by arrows. Densitometric scanning of the autoradiogram allowed us to calculate the percentage of interference with the following formula: (% unbound experimental − % unbound average) divided by (100% − % unbound average) (Ryder *et al.*, 1985). % unbound average indicates the percentage of DNA that is non-specifically excluded from the bound fraction and is calculated by taking the average of 10 bands outside the region protected against DNase I. % unbound experimental is calculated separately for every G residue (contained within the DNase I protected region).

TGGA/CN$_5$GCCAA (Gronostajski *et al.*, 1985). Site directed mutagenesis has revealed that alteration of this consensus sequence impairs or prevents binding of NFI (De Vries *et al.*, 1985; Schneider *et al.*, 1986). Sequences just outside the minimal recognition sequence modulate the binding affinity (De Vries *et al.*, 1985; Leegwater *et al.*, 1985).

To understand the interaction of NFI with its binding site in more detail we have resolved the positions on the DNA helix that are in close contact with NFI using methylation protection and alkylation interference assays. To study contacts between thymines and protein, we used 5-bromodeoxyuridine (5-BrdU) substituted DNA, prepared *in vitro*, as a photochemical probe. The results show that nuclear factor I binds almost exclusively at one side of the helix.

Results

The DNase I footprint of nuclear factor I is symmetrical

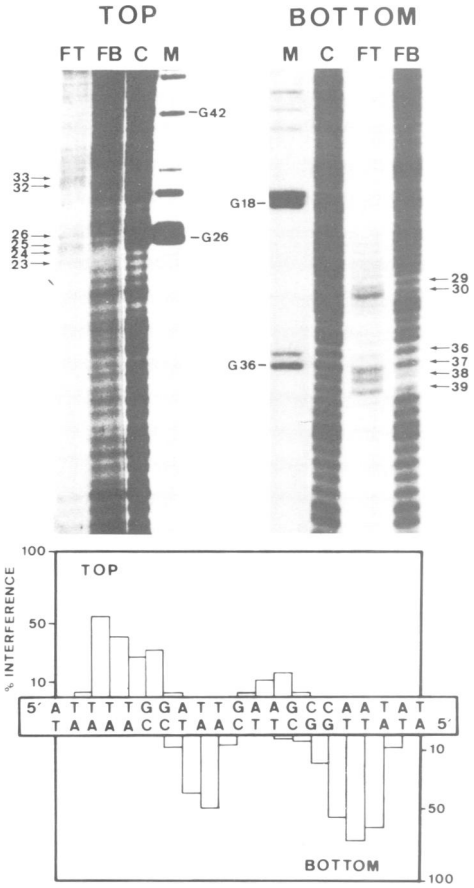
The consensus recognition sequence for binding of NFI shows partial dyad symmetry. This symmetry is even more pronounced in the binding site on the Ad2 origin. Several reports described the region protected against DNase I cleavage by NFI (Nagata *et al.*, 1983; Rawlins *et al.*, 1984; Leegwater *et al.*, 1985), but some discrepancies exist regarding the exact length and position of the protected region. These differences may arise from the use of different markers or from gel artifacts. Therefore we used several marker controls to establish the exact position. Using end-labeled binding sites synthesized *in vitro* (Figure 1A) we observed that NFI protects 24 phosphate bonds in both strands against DNase I cleavage (Figure 1B and C). In the top strand phosphate bonds 20 (the bond on the 3′-side of nucleotide 20) to 43 are protected and in the bottom strand phosphate bonds 42 (3′-side of nucleotide 42) to 19. Therefore, the protected regions are symmetrical with respect to the two-fold axis of symmetry drawn through the GC base pair at position 31. At the 5′-side of this base pair 11 bonds are protected in both strands and at the 3′-side 13 bonds.

NFI contactpoints within the Ad2 origin binding site

We employed chemical modification of DNA by alkylating reagents like dimethylsulphate and ethylnitrosourea to identify contactpoints of NFI with its recognition sequence.

Methylation protection

Dimethylsulphate methylates guanine residues at the N7 position in the major groove and adenine residues at the N3 position in the minor groove of the DNA helix. By subsequent chemical treatment the DNA backbone can be specifically cleaved either



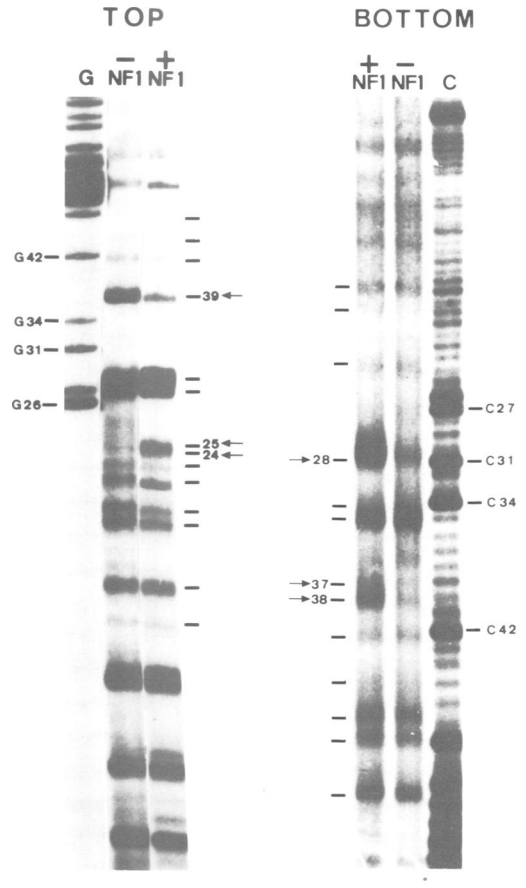


Fig. 4. Ethylation interference of NFI binding. For the top strand the *ClaI–PvuII* fragment of XD7 (5′-end-labeled at *ClaI* site) was used. The fragment we used for the bottom strand is described in the legend to Figure 2. Partially ethylated DNA was incubated with NFI and after filtration over nitrocellulose, protein bound (**FB**) and unbound (**FT**) fractions were analyzed on an 8% polyacrylamide gel. Interference is indicated by arrows and quantitatively represented in the histogram using the calculations outlined in the legend of Figure 3. The same numbering of phosphate bonds as described in the legend to Figure 1 was used.

Fig. 5. Determination of NFI contactpoints by u.v.-induced cleavage at BrdU residues. Fragments were labeled as described in the legend to Figure 1 except that 5-BrdUTP was included in the elongation reaction. After incubation of the 5-BrdUTP substituted DNA with NFI, the mixtures were irradiated with u.v. The products were separated on an 8% polyacrylamide gel. A chemical sequencing G or C reaction of the same fragment was used as marker. Due to a different mechanism of strand cleavage, marker and sample pattern are displaced about 2 bp. The positions of the T(BrDU) residues within the footprint area are indicated by bars. Since u.v.-induced breakage is not very homogeneous, some bands are hardly visible. Positions at which breakage is affected by NFI are indicated by arrows. Bands at positions 37 and 38 in the bottom strand and positions 24 and 25 in the top strand are not very well separated. The reason for these compressions is unknown.

at methylated G's alone or at both methylated G's and A's. As dimethylsulphate is a small molecule, protection from methylation of a certain G or A by a bound protein is a strong indication for intimate contacts between protein and DNA at that position.

Purified NFI was incubated at 0°C with an end-labeled fragment containing the binding site on the Ad2 origin. This mixture was subjected to partial methylation and specifically cleaved at methylated G's or G's and A's. Polyacrylamide gel electrophoresis revealed that in the top strand the G's at positions 26, 27 and 34 are protected from methylation (Figure 2, lane 2). Protection of A's does not occur (lane 4). No hypermethylated A's or G's can be observed. Similar experiments using the bottom strand show that G-residues at positions 35 and 36 are protected in the presence of NFI (Figure 2, lane 6).

Methylation interference

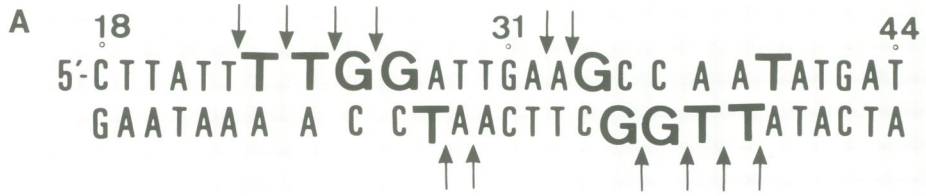
Another approach to identify contacts between protein and DNA makes use of previously modified DNA fragments which are subsequently incubated with an excess of the sequence-specific DNA-binding protein. Bound and unbound DNA fragments can be separated in a number of ways and analyzed on gel after specific strand cleavage at the modified nucleotide. Modification of the DNA at a position that normally interacts with a bound protein might interfere with binding to this protein. This results in the enrichment of the corresponding band in the unbound fraction and a depletion in the protein-bound fraction.

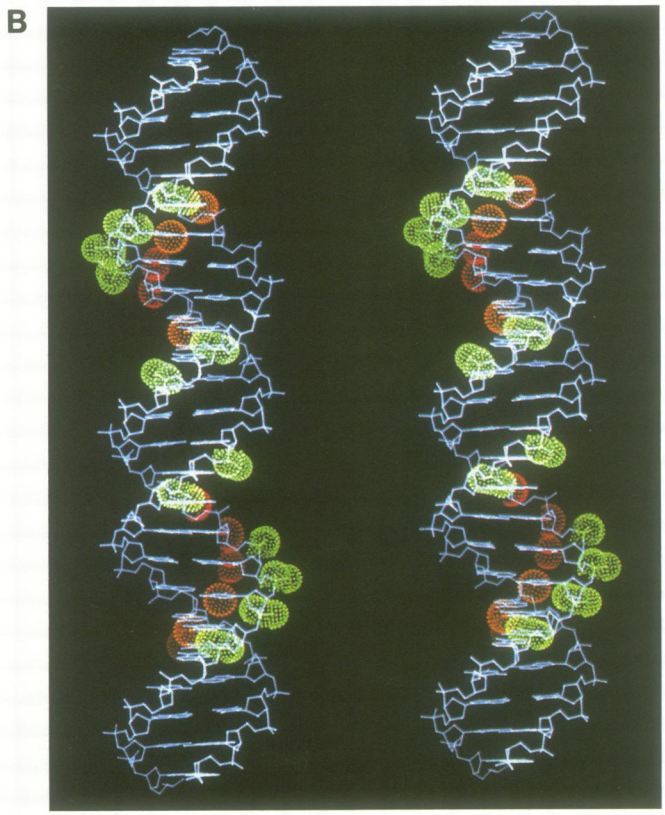
After incubation of NFI with partially methylated DNA at 0°C we separated protein bound and unbound fragments by filtration over nitrocellulose and analyzed both fractions on gel after cleavage at the modified nucleotide (Figure 3). Methylation of G's at positions 26, 27 and 34 in the top strand and positions 35 and 36 in the bottom strand appears to interfere with binding to NFI as the corresponding bands at these positions are predominantly in the filtrate fraction. The percentage of interference is plotted in a histogram (Figure 3).

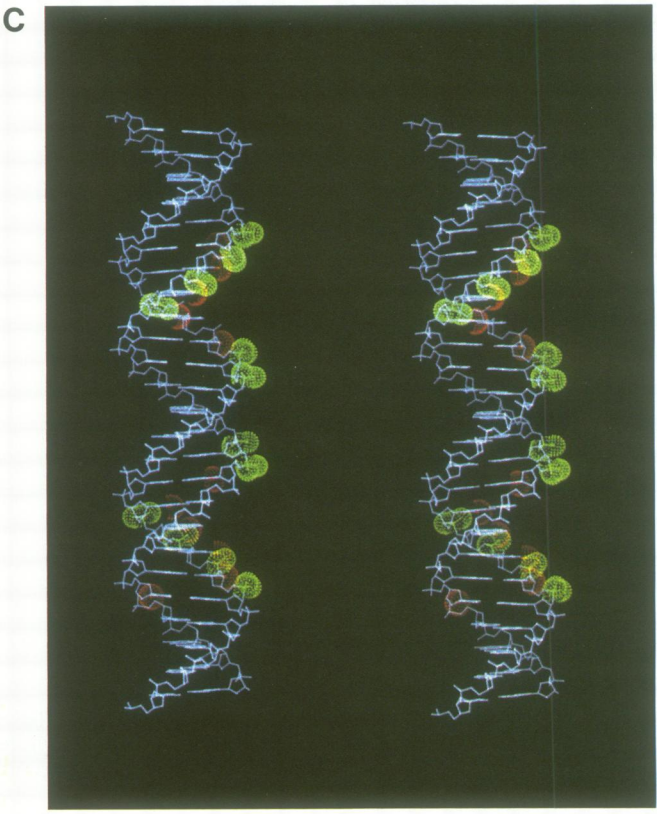
We conclude that both methylation protection and methylation interference identify the same guanine-N7 positions as contact points for NFI. The contacts which we detected in this way are limited to the two conserved blocks in the consensus recognition sequence. Moreover, only contacts with the major groove of the DNA have been detected.

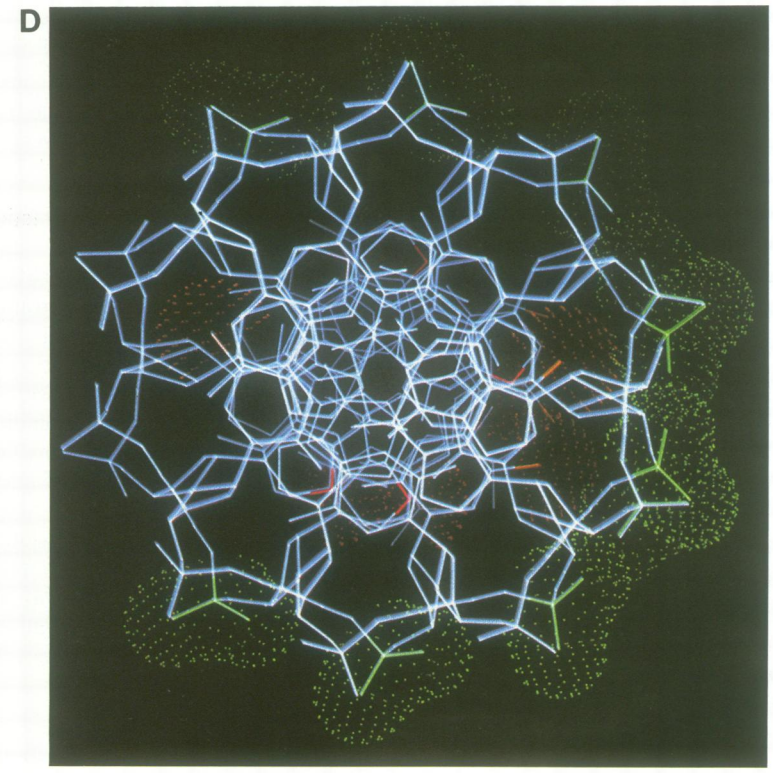
Ethylation interference

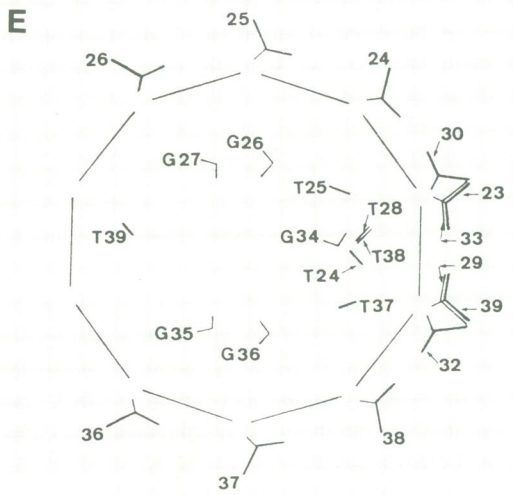
To detect contacts between NFI and the DNA backbone,


A
18
31
44
5'-CTTATTTTGGATTGAAGCCAATATGAT
GAATAAAACCTAACTTCGGTTATACTA


B


C


D


E
25
26
24
30
G27
G26
23
T25
T28
T39
G34
33
T38
T24
29
T37
39
G35
G36
32
36
38
37

phosphate groups were ethylated by ethylnitrosourea under conditions in which less than one phosphate per 200 bp is modified. After binding to NFI and filtration over nitrocellulose the DNA was cleaved at the sites of ethylation and separated using gel electrophoresis (Figure 4). Bands that are predominantly in the filtrate are indicated by arrows. Ethylation of the DNA at these positions impairs binding to NFI. This can either be caused by steric hindrance due to the ethyl group or by the removal of the negative charge of the phosphate. In both cases a close interaction between the protein and the phosphate is required to produce the observed effect. The histogram showing the percentage of interference indicates that some phosphates display only weak interference when ethylated whereas ethylation at other phosphates almost completely prevents binding to NFI. The strong dyad symmetry towards the axis through base pair 31 is also observed for phosphate contacts.

Analysis of T-contacts using 5-BrdU-substituted DNA

A less frequently employed technique developed to study contacts between protein and DNA makes use of DNA in which T-residues are substituted by 5-BrdU (Ogata and Gilbert, 1977). This is an analogue of thymine that carries a bromine group at the C5 position in the major groove instead of a methyl group. U.v.-irradiation of BrdU substituted DNA displaces the bromine, leaving a free radical at C5 that subsequently causes strand breakage at the sugar of the 5′-neighbouring nucleotide. Two different effects have been observed when a sequence-specific DNA-binding protein interacts closely with a bromine group. A number of amino acid side chains are able to neutralize the free radical by donating a hydrogen atom or crosslinking to the radical. This results in the loss of a band from the breakdown pattern (Ogata and Gilbert, 1977; Simpson, 1980, 1979). Alternatively, a bound protein can cause enhanced strand breakage at certain BrdU residues (Simpson, 1979). This can be explained by local structural or electrostatic changes in the DNA that might be induced by a bound protein that contacts the bromine. Electrostatic changes can alter the sensitivity of the bromine for ultraviolet absorption while structural changes can alter the position of the radical with respect to the sugar from which it seizes its hydrogen.

So far, DNA substituted *in vivo* with BrdU has been used in these experiments. We avoided this laborious procedure and adapted the M13 sequencing system to the synthesis of *in vitro* substituted DNA. Hybridization of 5′-end-labeled oligonucleotides T and B to ssM13 clones containing top and bottom strand of the Ad2 origin, respectively and subsequent primer extension was carried out as shown in Figure 1, except that 5-BrdUTP was included in the elongation reaction. Under appropriate 5-BrdUTP to TTP ratios DNA partially substituted in one strand is obtained. We established that incorporation of various amounts of BrdU in one strand, up to complete substitution of T-residues by BrdU, did not alter the footprint pattern (results not shown). This indicates that BrdU substitution had no effect on the binding of NFI although slight changes in binding affinity can not be excluded using this method.

Figure 5 shows that the presence of nuclear factor I causes a decreased breakage at the BrdU-residue at position 39 in the top strand. At positions 24 and 25 in the top strand and positions 37 and 38 in the bottom strand an increased strand breakage is observed. Cleavage at the BrdU-residue at position 28 in the bottom strand is strongly enhanced. This nucleotide is placed symmetrical to the G at position 34 in the other strand which was already shown to be contacted by NFI.

Most contacts are accessible from one side of the helix

In Figure 6A we have indicated all contacts that were identified. The overall symmetry towards the axis through base pair 31 is obvious. The spatial pattern of the contacts identified in the NFI binding site on the Ad2 origin is shown in computer-generated stereographics. Figure 6B shows a side view of the helix. Contactpoints are indicated as red (N7 of G residue), orange (methyl group on C5 of T residue) and green (phosphate) spheres. It is clear that 22 out of 23 contacts are accessible from one side of the helix. Only the T contact at position 39 on the top strand is located at the back side of the helix. This can more easily be seen in Figure 6C, where the helix is rotated 90° along its axis. A view down the axis of the helix (Figure 6D) again shows the distribution of contacts along one side of the helix. In the accompanying scheme the contacts are numbered to show that most contacts of the right half side of the recognition site (32−39) are in the lower right part of the top view whereas all contacts of the left half side (23−30) are positioned in the upper right part. Besides this small rotation of the two contacted blocks with respect to each other it is also apparent that the contacted T at position 39 is located opposite to the other contacts of this half of the recognition site.

An inverted NFI site is functional

As a consequence of the observed dyad symmetry in contactpoints one might expect that inversion of an NFI site is possible without loss of function. To test this hypothesis we constructed a mutant that contains an inverted NFI binding site and compared its efficiency of replication in a reconstituted system with a wild-type origin containing template. The sequences of the two template molecules are shown on top of Figure 7. pPL40 contains the first 40 bp of the left end of Ad2. We demonstrated previously that the first 40 bp are sufficient to support maximal NFI dependent replication *in vitro* (Leegwater *et al.*, 1985). Besides an intact NFI binding site, base pairs 9 to 18 are required for replication *in vitro* (Tamanoi and Stillman, 1983; Challberg and Rawlins, 1984). In pBIF the region spanning base pairs 22−40 is inverted with base pair 31 as the center of rotation. This stretch of DNA contains all positions identified to make contact with NFI and was able to bind NFI like wild-type (not shown). pPL40 and pBIF were digested by *Eco*RI and *Ava*I. The smaller fragment contains the origin of replication at the end of the molecule allowing it to serve as template in the *in vitro* system (Tamanoi and Stillman, 1982; Van Bergen *et al.*, 1983). Initiation of replication takes place by covalent attachment of a dCMP residue to the pre-terminal protein (Challberg *et al.*, 1980). This results in a decreased migration velocity of newly synthesized DNA. Addition of NFI to the reaction mixture enhances the replication reaction about 30-fold (De Vries *et al.*, 1985). From the results presented in Figure 7 we conclude that inversion of the NFI binding site does not have an effect on the magnitude of stimulation.

Fig. 6. Computer stereographics of the NFI binding site. In (**A**) all NFI contacts which we identified are indicated with arrows or in bold type. (**B**) shows a side-view of the NFI binding site. The Van der Waals radii of GN_7-atoms (red), TM_5-groups (orange) and phosphates (green), that are contacted by NFI, are displayed as coloured spheres. In (**C**) the helix is turned 90° rightwards along its axis and in (**D**) the helix of (C) is tilted 90° to obtain a view down the axis of the helix. Atoms in front are displayed brighter than atoms more backwards. Base pair 43 is the front base pair of (D) and the bottom base pair of (B) and (C). (B) and (C) may be viewed in a stereoscope. In (E) the various contacts of the top-view are numbered.

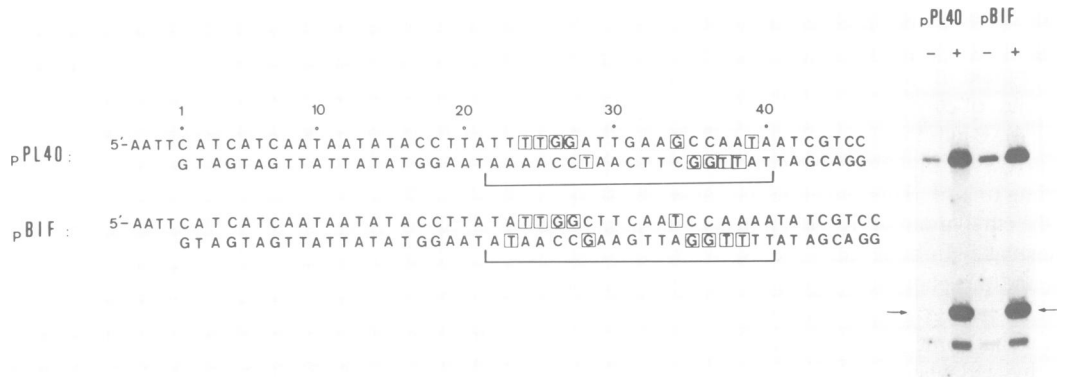


Fig. 7. *In vitro* replication with a template containing an inverted NFI binding site. The wild-type plasmid (pPL40) and the inversion mutant plasmid (pBIF) were linearized by *Eco*RI digestion as indicated in the sequence. Numbering is the same as in Figure 1. Base pairs that are contacted by NFI in wild-type are boxed. The region within the bracket has been inverted around base pair 31. The *in vitro* replication reaction has been performed in the presence (+) or the absence (−) of NFI and the products were separated on an SDS containing agarose gel (Van Bergen *et al.*, 1983). The band representing replication (see text) is indicated by an arrow.

Discussion

We have performed a detailed contactpoint analysis of the human HeLa-protein NFI. Several features of NFI binding are clearly visualized by the computer-generated stereoscopic representations shown in Figure 6. It is obvious that almost all contactpoints are accessible from one face of the DNA helix. The contacts identified are made within two successive turns of the major groove and with the phosphate backbone. One should keep in mind, however, that in the minor groove only contacts with adenines can be studied and that other minor groove contacts may still exist. The strong dyad symmetry towards the axis through base pair 31 that is apparent from the DNA sequence is also present on the level of contactpoints and DNase I footprinting.

With regard to the spatial distribution and symmetry of contactpoints there is striking similarity with a number of well-defined recognition sites of prokaryotic repressor and/or activator proteins like CAP (Simpson, 1980), phage λ-repressor (Pabo *et al.*, 1982), and cro protein (Anderson *et al.*, 1981). These recognition sites also consist of two blocks of about 5 bp, which show dyad symmetry and are separated by a spacer of 4−6 bp long. Crystallography and n.m.r.-studies provided detailed information about the structure of these proteins (for a review, see Takeda *et al.*, 1983). Model building and, more recently, the co-crystallization of the phage 434 repressor with its binding site (Anderson *et al.*, 1985) revealed the mode of binding in a number of cases. Binding takes place by dimers or tetramers whereby both symmetrical blocks are contacted by one monomer. The contacting part of the protein has a characteristic conserved structure that consists of two consecutive α-helices. One α-helix protrudes into the major groove making sequence specific contacts with the conserved block whereas the other α-helix lies across the major groove and is supposed to make contacts primarily with the phosphate backbone. The obtained image of contacts on the NF1 binding site is well shaped to accommodate a similar array of protein binding.

So far only two eukaryotic sequence specific DNA-binding proteins have been examined in detail for contacts with their recognition sites. The HeLa promoter specific transcription factor SP1 was shown to have major groove contacts on only one side of the DNA helix. In this case the core recognition sequence consists of only 6 bp and is highly asymmetric which is reflected in the observed contactpoints (Gidoni *et al.*, 1984). The NFI binding site differs from the SP1 site in these respects and shows more similarity to the SV40 large T-antigen binding sites. Large T-antigen was shown to bind as a dimer to a sequence consisting of two identical pentanucleotides separated by a spacer of seven nucleotides length (Jones and Tjian, 1984). Major groove contacts are all located at one side of the helix whereas large T-antigen contacts phosphates on either side of the helix. Possibly T-antigen possesses 'arms' that wrap around the helix and contact the phosphate backbone at the backface. Backface contacts have previously been identified for λ-repressor and λ-cro protein, but the back side contact for the 'repressor-arm' is located in the major groove (Pabo *et al.*, 1982) and the 'cro-arm' is supposed to contact the minor groove (Ohlendorf *et al.*, 1982). We identified only one NF1 contact on the back side, the methyl group at the C5 position of the T at 39 which is clearly separated from the other five contacts in this turn of the major groove (see Figure 6). Evidence that this contact is important for binding comes from deletion mutant studies that indicate that the base pairs at position 39 and possibly position 40 are essential for binding of NFI (Leegwater *et al.*, 1985). The question whether NFI possesses an 'arm' that is responsible for this contact can only be resolved by determination of protein structure. The lack of identification of a contact at position 23 of the binding site is most likely due to the asymmetry in sequence at this position (A instead of T). A remarkable feature of the T-antigen binding site is the possibility to invert one of the two pentanucleotides resulting in a direct repeat of the two pentanucleotides which has a slightly higher binding affinity (Ryder *et al.*, 1985). We think it unlikely that such an inversion is possible in case of NFI since all NFI binding sites identified so far are highly palindromic. There are also strong sequence requirements for the 7 bp spacer region to allow efficient T-antigen binding which probably reflects the need of a certain spacer structure (Ryder *et al.*, 1985). Between the different NFI sites this spacer region is highly degenerated. We conclude that there are marked structural differences in binding of NFI, SPI and large T-antigen.

We made use of labeled binding sites produced by extension of a 5′-end-labeled oligonucleotide instead of a double-stranded DNA fragment purified from polyacrylamide gels. This procedure offers several advantages like speed, versatility and labeling to high specific activity. An additional advantage is the possibility to incorporate modified nucleotides. BrdU substituted DNA can thus be made to any desired substitution grade. Another nucleotide analogue that we employed for contactpoint analysis

was 5-bromodeoxycytidine triphosphate. We obtained 5-BrdC substituted DNA that could be specifically cleaved by u.v. irradiation, but NFI bound only weakly to this DNA even at low substitution grades. Surprisingly this effect could not be attributed to any specific C residue in the NFI binding site. Currently we are investigating this observation more extensively.

Little is known about the mechanism of NFI stimulation. We have demonstrated that at least in the Ad2 origin the recognition site can be inverted without any effect on the stimulation of replication. Recently it was shown that a similar inversion mutant supported pTP-dCMP complex formation at a level that was ~60% of wild-type. This small decrease, which we did not observe, could be due to the base substitutions at positions 19 and 20 in that particular mutant (Adhya *et al.*, 1986). The observation that inverted NFI sites are fully functional could be explained easily by binding of NFI as a dimer or a tetramer to its symmetrical binding site and makes other possible modes of binding less likely. However, it should be kept in mind that inversion of highly asymmetric binding sites like the SP1 binding sites in the SV40 transcriptional control region also has no effect on the stimulation of transcription (Barrera-Saldana *et al.*, 1985) showing that binding of an asymmetric protein in one direction can still cause an equal action in both directions. Recent studies have shown that a number of eukaryotic gene control regions contain multiple adjacent binding sites for one or several different sequence specific DNA binding proteins, suggesting the involvement of large nucleoprotein structures in initiation of transcription (Takahashi *et al.*, 1986; Topol *et al.*, 1985; Heberlein *et al.*, 1985; Jones *et al.*, 1985; Jones and Tjian, 1985). NFI function might also be dependent on such structures. Bordering the NFI binding site are the recognition sites for the pTPpol complex (base pairs 9–18, Rijnders *et al.*, 1983) and the binding site of another recently identified HeLa sequence specific DNA-binding protein, nuclear factor III (Pruijn *et al.*, 1986). The location of NFI on one side of the DNA helix might be an important factor in proper functioning. This idea is supported by the observation that insertions or deletions between the NFI binding site and the pTPpol recognition site have a strong negative effect on NFI stimulation of replication (Adhya *et al.*, 1986) while the magnitude of the effect is influenced by the exact length and sequence of the insertion (De Vries and Van der Vliet, unpublished). Such insertions will result in relative twisting of the binding sites which will complicate protein–protein interactions. A similar effect has recently been observed in a detailed analysis of the SV40 transcriptional control region (Takahashi *et al.*, 1986). Optimal early transcription is dependent on at least three different sequence elements: a TATA-box, six SP1-binding sites and the SV40 enhancer. These elements have to be aligned in a stereospecific manner to allow their respective binding proteins to be active. This alignment is disrupted by inserting odd multiples of half a helix-turn in between the different binding sites, resulting in a strong decrease of transcription. Insertion of even multiples of half a helix-turn had only a moderate effect. Based on the results presented in this paper we conclude that the adenovirus *in vitro* DNA replication system is an attractive model system for further investigation of the importance of stereospecific alignment of regulatory DNA binding proteins and the possible mode of interaction of NFI with other proteins.

Materials and methods

Recombinant DNA plasmids

XD7 was constructed by cloning the 1338-bp *Xba*I E-fragment of Ad2 in the *Eco*RI site of pBR322 (Pearson *et al.*, 1983). In MXE-1 the *Eco*RI fragment of XD7 is inserted in the *Eco*RI site of M13mp8 (De Vries *et al.*, 1985). MXE-4 contains the same fragment inserted in the *Eco*RI site of M13mp9 in the reversed orientation. We constructed pPL73 by cloning the 73 bp *Eco*RI–*Fnu*DII fragment of XD7 in the *Eco*RI–*Hinc*II site of pUC12. pBIF was constructed by cloning complementary synthetic oligonucleotides (5′-TTATATTGGCTTCAATCCAAAA and 5′-TAATTTTGGATTGAAGCCAATA) in the *Sau*I–*Eco*RI RV site of Xpm2223 (De Vries *et al.*, 1985). The 5′-overhanging ends of the hybridized oligonucleotides were partially filled in by incubation with the large fragment of *E. coli* DNA polymerase I (30 min, 25°C, 1 μg DNA, 0.1 U polymerase). In this way part of the molecule becomes blunt-ended on one side, while the other side is still cohesive with the *Sau*I site. The sequence of the resulting recombinants was verified by DNA sequence analysis.

DNase I footprinting

NFI used for footprinting, contactpoint analysis and *in vitro* DNA replication was purified as described (Leegwater *et al.*, 1985). Synthetic oligonucleotides T (5′-GTAAACGACGGCCAT) or B (5′-ACGCCCCGCGCCACGTC) were 5′-end labeled by incubating 1.5 pmol oligonucleotide for 45 min at 37°C in a reaction volume of 20 μl containing 5 units T4 polynucleotide kinase (Boehringer), 25 pmol [γ-^{32}P]ATP (sp. act. >5000 Ci/mmol, Amersham), 50 mM Tris–HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM dithiothreitol. T4 polynucleotide kinase was inactivated during 15 min at 68°C. 5 μl labeled oligonucleotide T or B was added to 1.5 μg ssMXE-4 or 1.5 μg 22MXE-1 and adjusted to 50 mM NaCl, 66 mM Tris–HCl (pH 7.5), 6.6 mM $MgCl_2$, 2.5 mM dithiothreitol (final volume 10 μl), heated for 3 min at 80°C and allowed to hybridize for 1 h at room temperature. Primer extension was carried out by addition of 5 U *E. coli* DNA polymerase I (Klenow fragment, Boehringer) after adjusting the reaction mixture to 50 μM ATP, 150 μM dNTP's, 50 mM NaCl, 66 mM Tris–HCl (pH 7.5), 6.6 mM dithiothreitol (final volume 50 μl). After 2 h at 30°C [α-^{32}P]-ATP was removed by filtration over a 1 ml Sepharose CL-4B column. The yield of this procedure was about 3×10^6 c.p.m. (Cerenkov), incorporated into DNA that has been elongated at least 400 bp as determined by restriction enzyme digestion. Labeled DNA (10 000 c.p.m.) was incubated with 0.5 mU (Leegwater *et al.*, 1985) NFI for 60 min at 0°C in a buffer containing 25 mM Hepes–NaOH (pH 7.5), 5 mM $MgCl_2$, 4 mM dithiothreitol, 150 mM NaCl and 50 ng pBR322. The mixture was incubated for 90 s at 24°C with 0.1 U DNase I (Mol. biol. grade, Boehringer). Further handling and analysis of samples has been described previously (Leegwater *et al.*, 1985).

Contactpoint analyses

The 454 bp *Eco*RI–*Pvu*II fragment of XD7 [5′-end-labeled (Maxam and Gilbert, 1980) at the *Eco*RI site] and the 270 bp *Hin*dIII–*Pvu*I fragment of pPL 73 (5′-labeled at the *Hin*dIII site) were used to analyze the top and bottom strand, respectively. For the ethylation interference assay the *Cla*I–*Pvu*II fragment of XD7 (5′-labeled at *Cla*I site) was used for the top strand. For the interference assays DNA templates were partially methylated with dimethylsulphate or partially ethylated with ethylnitrosourea as described (Siebenlist and Gilbert, 1980). The modified DNA (25–50 fmol, 30 000 c.p.m.) was incubated with 1.0 mU NFI for 60 min at 0°C in a volume of 50 μl containing 25 mM Hepes–NaOH (pH 7.5), 5 mM $MgCl_2$, 5 mM DTT, 100 mM NaCl and 200 ng pBR322. Subsequently the mixture was passed over a nitrocellulose filter (Millipore, HAWP). Filters were washed three times with 50 μl of the same buffer. Filterbound material was eluted in 0.5 M NH_4Ac, 0.1% SDS. Further handling of the fractions has been described (Siebenlist and Gilbert, 1980). The products were analysed on 8 or 10% denaturing polyacrylamide gels. For the methylation protection assay, 5′ end-labeled DNA was incubated with NFI under the conditions described above After 1 h, protection against methylation by dimethylsulphate was determined as described (Siebenlist and Gilbert, 1980). 5-BrdU substituted DNA was synthesized *in vitro* by the above described primer extension procedure. The only modification was the addition of 150 μM 5-BrdUTP and 50 μm dTTP, instead of 150 μM, to the elongation reaction. Assuming that 25% of the bromine residues are cleaved by u.v. (as estimated from breakdown of fully substituted DNA) this leads to a 20% BrdU substitution level. The substituted DNA (30 000 c.p.m.) was incubated at 0°C with 1.0 mU NFI in a buffer containing 150 mM NaCl, 25 mM Hepes–NaOH (pH 7.5), 5 mM $MgCl_2$, 4 mM DTT, 50 ng pBR322, 2.5 μg BSA (final volume 100 μl). After 1 h the mixture was irradiated with u.v.-light for 2 min by placing it at 7.5 cm distance from a Philips SP500 mercury lamp. The sample holder was jacketed by a cooled, circulating solution of (Whatmann 1 filtered) 0.2 M $CuSO_4$, 1.0 M $NiSO_4$, through which only transmission of light with a wavelength around 320 nm (the absorption maximum of 5-BrdU) takes place. The sample was phenol extracted, alcohol precipitated and analyzed on an 8% denaturing polyacrylamide gel.

Computer graphics

The three dimensional projections of the NFI site were constructed at the facilities of the Computer Graphics group at EMBL, Heidelberg (Dr Heinz Bosshard). Base pairs 20 to 43 of Ad2 are visualized according to the co-ordinates of idealized B-DNA and Van der Waals surfaces of the contactpoints are displayed interactively

as dotted spheres, using the UCSF MIDAS program (Langridge *et al.*, 1981; Bash *et al.*, 1983), on an Evans and Sutherland Picture System.

Other methods

DNA replication *in vitro* was studied using linearized plasmid DNA and purified pTP-pol, DBP, NFI as described in detail previously (De Vries *et al.*, 1985).

Acknowledgements

We would like to thank Drs J.Postma, D.Banner, and H.Bosshard of the EMBL, Heidelberg, for their help with computer graphics. We also thank Dr J.Van Boom for providing synthetic oligonucleotides. This work was supported in part by the Netherlands Foundation for Chemical Research (SON) with financial aid from The Netherlands Organization for the Advancement of Pure Research (ZWO).

References

Adhya,S., Shneidman,P.S. and Hurwitz,J. (1986) *J. Biol. Chem.*, **261**, 3339–3346.

Anderson,J.E., Ptashne,M. and Harrison,S.C. (1985) *Nature*, **316**, 596–601.

Anderson,W.F., Ohlendorf,D.H., Takeda,Y. and Matthews,B.W. (1981) *Nature*, **290**, 754–758.

Barrera-Saldana,H., Takahashi,K., Vigneron,M., Wildeman,A., Davidson,I. and Chambon,P. (1985) *EMBO J.*, **4**, 3839–3849.

Bash,P., Pattabiraman,N., Huang,C., Ferrin,T. and Langridge,R. (1983) *Science*, **222**, 1325.

Borgmeyer,U., Nowock,J. and Sippel,A.E. (1984) *Nucleic Acids Res.*, **12**, 4295–4311.

Cameron,V. and Uhlenbeck,O.C. (1977) *Biochemistry*, **16**, 5120–5126.

Challberg,M.D., Desiderio,S.V. and Kelly,T.J.,Jr (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5105–5109.

Challberg,M.D. and Rawlins,D.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 100–104.

De Vries,E., Van Driel,W., Tromp,M., Van Boom,J. and Van der Vliet,P.C. (1985) *Nucleic Acids Res.*, **13**, 4935–4952.

Gidoni,D., Dynan,W.S. and Tjian,R. (1984) *Nature*, **312**, 409–413.

Gronostajski,R.M., Nagata,K. and Hurwitz,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4013–4017.

Gronostajski,R.M., Adhya,S., Nagata,K., Guggenheimer,R.A. and Hurwitz,J. (1985) *Mol. Cell. Biol.*, **5**, 964–971.

Guggenheimer,R.A., Stillman,B.W., Nagata,K., Tamanoi,F. and Hurwitz,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3069–3073.

Hay,R.T. (1985) *EMBO J.*, **4**, 421–426.

Heberlein,U., England,B. and Tjian,R. (1985) *Cell*, **41**, 965–977.

Hennighausen,L., Siebenlist,U., Danner,D., Leder,P., Rawlins,D., Rosenfeld,P. and Kelly,T.,Jr (1985) *Nature*, **314**, 289–292.

Jones,K.A. and Tjian,R. (1984) *Cell*, **36**, 155–162.

Jones,K.A., Yamamoto,K.R. and Tjian,R. (1985) *Cell*, **42**, 559–572.

Jones,K.A. and Tjian,R. (1985) *Nature*, **317**, 179–182.

Langridge,R., Ferrin,T.E., Kuntz,I.D. and Connolly,M.L. (1981) *Science*, **211**, 661–666.

Leegwater,P.A.J., Van Driel,W. and Van der Vliet,P.C. (1985) *EMBO J.*, **4**, 1515–1521.

Leegwater,P.A.J., Van der Vliet,P.C., Rupp,R.A.W., Nowock,J. and Sippel,A.E. (1986) *EMBO J.*, **5**, 381–386.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.

Nagata,K., Guggenheimer,R.A., Enomoto,T., Lichy,J.H. and Hurwitz,J. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6438–6442.

Nagata,K., Guggenheimer,R.A. and Hurwitz,J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6177–6181.

Ohlendorf,D.H., Anderson,W.F., Fisher,R.G., Takeda,Y. and Matthews,B.W. (1982) *Nature*, **298**, 718–722.

Ogata,R. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4973–4976.

Pabo,C.O., Krovatin,W., Jeffrey,A. and Sauer,R.T. (1982) *Nature*, **298**, 441–443.

Pearson,G.D., Chow,K.-C., Enns,R.E., Ahern,K.G., Corden,J.L. and Harpst, J.A. (1983) *Gene*, **23**, 293–305.

Pruijn,G.J.M., Van Driel,W. and Van der Vliet,P.C. (1986) *Nature*, **322**, 656–659.

Rawlins,D.R., Rosenfeld,P.J., Wides,R.J., Challberg,M.D. and Kelly,T.J.,Jr (1984) *Cell*, **37**, 309–319.

Rijnders,A.W.M., Van Bergen,B.G.M., Van der Vliet,P.C. and Sussenbach,J.S. (1983) *Nucleic Acids Res.*, **11**, 8777–8789.

Rosenfeld,P.J. and Kelly,T.J. (1986) *J. Biol. Chem.*, **261**, 1398–1408.

Ryder,K., Vakalopoulou,E., Mertz,R., Mastrangelo,J., Hough,P., Tegtmeyer,P. and Fanning,E. (1985) *Cell*, **42**, 539–598.

Schneider,R., Gander,I., Muller,U., Mertz,R. and Winnacker,E.L. (1986) *Nucleic Acids Res.*, **14**, 1303–1317.

Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 122–126.

Siebenlist,U.L., Henninghausen,L., Battey,J. and Leder,P. (1984) *Cell*, **37**, 381–391.

Simpson,R.B. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3233–3237.

Simpson,R.B. (1980) *Nucleic Acids Res.*, **8**, 759–766.

Sippel,A.E., Fritton,H., Theisen,M., Borgmeyer,U., Strech-Jurk,U., Igo-Kemenes,T. (1986) In Grodzicker,T., Sharp,P.A. and Botchan,M.R. (eds), *Cancer Cells 4: DNA Tumor Viruses: Gene Expression and Replication*. Cold Spring Harbor Laboratory Press, NY, in press.

Takahashi,K., Vigeron,M., Matthes,H., Wildeman,A., Zenke,M. and Chambon,P. (1986) *Nature*, **319**, 121–126.

Takeda,Y., Ohlendorf,D.H., Anderson,W.F. and Matthews,B.W. (1983) *Science*, **22**, 1020–1026.

Tamanoi,F. and Stillman,B.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6446–6450.

Topol,J., Ruden,D.M. and Parker,C.S. (1985) *Cell*, **42**, 527–537.

Van Bergen,B.G.M., Van der Leij,P.A., Van Driel,W., Van Mansfeld,A.D.M. and Van der Vliet,P.C. (1985) *Nucleic Acids Res.*, **11**, 1975–1989.

Received on 10 October 1986; revised on 3 November 1986